

ABSTRACT

A language model generation and accumulation apparatus (10) that generates and accumulates language models for speech recognition is comprised of: a higher-level N-gram generation and accumulation unit (11) that generates and accumulates a higher-level N-gram language model obtained by modeling each of a plurality of texts as a string of words including a word string class having a specific linguistic property; and a lower class dependent word N-gram generation and accumulation unit (12) that generates and accumulates a lower-level N-gram language model obtained by modeling a sequence of words included in each word string class.

(19) 世界知的所有權機關
國際事務局



A standard linear barcode located at the bottom of the page, used for tracking and identification of the publication.

(43) 国際公開日
2004年4月22日(22.04.2004)

PCT

(10) 国際公開番号
WO 2004/034378 A1

(5) 國際特許分類⁷:

G101, 15/18

(72) 発明者: および

(二) 國際出版商名：

PCT/JP2003/012701

(25) 國際出願の言語.

日文語

(26) 國際公關の言語・

日本語

(30) 优先権データ:

特願2002-294500 2002年10月8日(08.10.2002) JP
特願2003-276844 2003年7月18日(18.07.2003) JP

(71) 出願人(米国を除く全ての指定国について): 松下電器産業株式会社 (MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.) [JP/JP]: 〒571-8501 大阪府門真市大字門真 1006 番地 (Osaka JP).

(72) 発明者: および

(75) 発明者/出願人(米国についてのみ): 沖本 純幸 (OKIMOTO, Yoshiyuki) [JP/JP]: 〒619-0224 京都府 相楽郡
木津町 兜台2丁目2番1-D-105号 Kyoto (JP). 遠藤 充
(ENDO, Mitsuru) [JP/JP]: 〒216-0015 神奈川県 川崎市
宮前区 菅生3丁目33番17-319号 Kanagawa (JP). 西崎 誠
(NISHIZAKI, Makoto) [JP/JP]: 〒177-0034 東京都 犀馬
区 富士見台1丁目21番3-101号 Tokyo (JP).

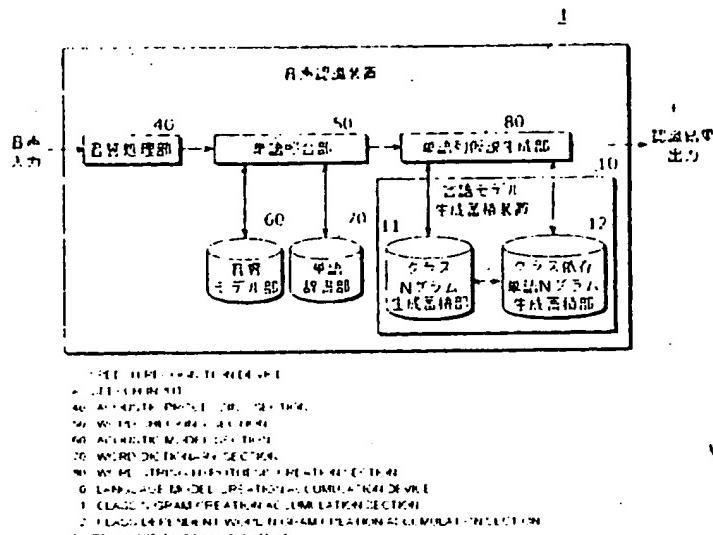
(74) 代理人: 新居 広守 (Nii,Hiromori); 〒532-0011 大阪府
大阪市淀川区西中島3丁目11番26号 新大阪末広セン
タービル3F 新居国際特許事務所内 Osaka (JP).

(81) 指定国(国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB,
BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK,
DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,

有蔓繞

(54) Title: LANGUAGE MODEL, CREATION/ACCUMULATION DEVICE, SPEECH RECOGNITION DEVICE, LANGUAGE MODEL, CREATION METHOD, AND SPEECH RECOGNITION METHOD

(54) 発明の名称: 言語モデル生成装置、音声認識装置、言語モデル生成方法および音声認識方法



WO 2004/034378

(57) Abstract: A language model creation/accumulation device (10) for creating and accumulating a language model for speech recognition includes: an upper node class N-gram creation/accumulation section (11) for creating and accumulating an upper node N-gram language model obtained by modeling a plurality of texts as a word string containing a word string cluster having a particular language characteristic; and a lower node class-dependent word N-gram creation/accumulation section (12) for creating and accumulating a lower node N-gram language model obtained by modeling a word string in a word string cluster.

(57) 要約：音声認識用の言語モデルを生成し、蓄積する言語モデル生成蓄積装置（10）は、複数のテキストを、特定の言語的特性を有する単語列クラスを含む単語の系列としてモデル化した上位Nグラム言語モデルを生成し、蓄積する上位のクラスNグラム生成蓄積部（11）と、単語列クラス内の単語の系列をモデル化した下位Nグラム言語モデルを生成し、蓄積する下位のクラスNグラム生成蓄積部（12）。

有葉綱



LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NL, NO, NZ, OM, PG, PL, PT, RO, RU, SC, SD, SE,
SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, YU, ZA, ZM, ZW.

(84) 指定国(広域): ARIPO特許(GH, GM, KE, LS, MW, MZ,
SD, SL, SZ, TZ, UG, ZM, ZW), ヨーラシア特許(AM,
AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許
(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB,
GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR).

OAPI 特許(BE, BJ, CE, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

添付公開書類:

一 國際調査報告書

2文字コード及び他の略語については、定期発行される
各PCTガゼットの巻頭に掲載されている「コードと略語
のガイド」を参照。